**主题:**数据提取流程  
 高血压并发症风险预测

研究对象：高血压患者人群

预测目的：实现患严重并发症事件的风险预警

实验组：高血压患者并患严重并发症，如低血压、脑梗、心衰等

对照组：高血压患者，其他相对平稳，一般除去心血管、代谢、神经、眼部疾病、肾、肺等并发症状。

Oracle处理相关表：clinical\_icd\_bit\_htn0817：门诊高血压人群（3位ICD编码），TM\_PERSON\_INFO：居民信息表，TM\_DIMA\_HTN\_REG：高血压登记表，TM\_DIMA\_HTN\_FOLLOWUP：高血压随访表，TM\_OUTCLINICAL：门诊表等

任务流程：

1、加载数据至Oracle，该数据来自李奇提取出的病症和患病人群的0,1二维表，大致422951\*543维，该表是经过整理过得门诊病症表（病症划分为543种）。

2、关联步骤1加载后的表和高血压登记表，生成门诊高血压表 **Outclinical\_High\_Htn\_Reg**，这里表的人群就是面向高血压人群，包括门诊诊断的症状，也就是那543种症状。

**更改**：高血压登记表 应使用确诊日期唯一最新处理过的高血压登记表。

3、关联高血压登记表和高血压随访表，从确诊时间算起，筛选出只跟踪了3年以下且没有心血管、代谢、神经、眼部疾病、肾、肺等并发症的人群。

处理的具体过程：

* **首先筛选出3年以下的高血压人群MPI\_PERSON\_ID， 与Outclinical\_High\_Htn\_Reg链接操作后，去掉243中严重并发症的人群，得到三年内高血压健康人群的MPI\_PERSON\_ID，其中有大约3万多位患者，5万多条记录。**
* **在Outclinical\_High\_Htn\_Reg表中去掉三年内高血压健康人群的MPI\_PERSON\_ID，得到的表New\_Outclinical\_High\_Htn\_Reg中 有实验组，对照组。**
* **进一步提取对照组， 提取的过程和提取三年内的高血压健康人群一样。对照组有65492条记录，不同人群有48118**

4、由步骤2生成的表，去除步骤3含有的人群，所得到的人群就是我们研究的对象人群。

5、 由步骤4得到的对象人群中，筛选出实验组（如：患有脑梗或者心衰或者低血压等并发症）

* **筛选出脑梗、心衰类的ICD编码。 （？？？如何筛）**
* **在去除掉New\_Outclinical\_High\_Htn\_Reg表中的对照组，根据实验组的ICD编码进而提取对应的实验组 。**

6、 由步骤4得到的对象人群中，筛选出对照组，满足高血压登记表里的高血压并发症没有以及步骤4得到的对象人群中也没有并发症（如：含有心血管、代谢、神经、眼部疾病、肾、肺）。